Derivations of $\frac{\partial L}{\partial z}$

Sigmoid activation with Logistic loss

$$a(z) = \sigma(z) = \frac{1}{(1+e^{-z})}$$

$$L(a, y) = -(y \ln a + (1-y) \ln (1-a))$$

For those who are curious about where the $dz = a - y$ comes from.

If you're curious, here is the derivation for

$$\frac{\partial L}{\partial z} = a - y$$

By the Chain rule: $\frac{\partial L}{\partial z} = \frac{\partial L}{\partial a} \frac{\partial a}{\partial z}$

Step 1: Solve for $\frac{\partial L}{\partial a}$

$$L = -(y \ln a + (1-y) \ln (1-a))$$

take the derivative with respect to a. Remember that there is an additional -1 in the last term when we take the derivative of $log(1-a)$ with respect to a

$$\frac{\partial L}{\partial a} = -y \frac{1}{a} - (1-y) \frac{1}{(1-a)} (-1) = -y \frac{1}{a} + (1-y) \frac{1}{(1-a)}$$

$$\frac{\partial L}{\partial a} = \frac{-y}{a} + \frac{1-y}{1-a}$$

$$\frac{\partial L}{\partial a} = \frac{-y(1-a)}{a(1-a)} + \frac{a(1-y)}{a(1-a)}$$

$$\frac{\partial L}{\partial a} = \frac{-y + a\,y + a - a\,y}{a(1-a)}$$

$$\frac{\partial L}{\partial a} = \frac{a-y}{a(1-a)}$$

Step 2: Solve for $\frac{\partial a}{\partial z}$

$$\frac{\partial a}{\partial z} = \frac{d}{dz}\sigma(z)$$

The derivative of the sigmoid functions the form (not derived):

$$\frac{d}{dz}\sigma(z) = \sigma(z)\left(1 - \sigma(z)\right)$$

Recall that $\sigma(z) = a$, because we defined a, the activation, as the output of the sigmoid activation function. So we can substitute into the formula to get:

$$\frac{\partial a}{\partial z} = a\,(1 - a)$$

Step 3: $\frac{\partial L}{\partial z}$

Multiply step 1 and step 2 to get the result.

$$\frac{\partial L}{\partial z} = \frac{\partial L}{\partial a}\frac{\partial a}{\partial z}$$

From step 1: $\frac{\partial L}{\partial a} = \frac{a-y}{a\,(1-a)}$

From step 2: $\frac{\partial a}{\partial z} = a\,(1-a)$

$$\frac{\partial L}{\partial z} = \frac{a - y}{a\,(1-a)}\,a\,(1-a)$$

$$\frac{\partial L}{\partial z} = a - y$$

Sigmoid activation with MSE loss

$$a(z) = \sigma(z) = \frac{1}{(1 + e^{-z})}$$

$$L(a, y) = \frac{1}{2}\,(a - y)^2$$

$$\frac{\partial L}{\partial a} = a - y$$

$$\frac{\partial a}{\partial z} = a\,(1 - a)$$

$$\frac{\partial L}{\partial z} = \frac{\partial L}{\partial a}\frac{\partial a}{\partial z} = (a - y)\,(a\,(1 - a))$$

Linear activation with MSE loss

$$a(z) = z$$

$$L(a, y) = \frac{1}{2}\,(a - y)^2$$

$$\frac{\partial L}{\partial a} = a - y$$

$$\frac{\partial a}{\partial z} = 1$$

$$\frac{\partial L}{\partial z} = \frac{\partial L}{\partial a}\frac{\partial a}{\partial z} = a - y$$

Softmax activation with cross entropy loss function

$$a_i(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{N} e^{z_j}}$$

$$L(a, y) = -y \ln a$$

$$\frac{\partial L}{\partial a} = \frac{-y}{a}$$

$$\frac{\partial a}{\partial z} = \frac{a\,(y-a)}{y}$$

$$\frac{\partial L}{\partial z} = \frac{\partial L}{\partial a}\,\frac{\partial a}{\partial z} = a - y$$

LaTex Math Examples

$$LD = a - y$$

$$z = x + ya^2 + b^2 = c^2 \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

$$\frac{\partial u}{\partial t} = h^2 \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \right)$$

$$\lim_{x \to +\infty}, \inf_{x > s}$$

$$\sum_{k=1}^{n} k^2 = \frac{1}{2} n(n+1)$$

$$\int_a^b f(x)\, dx$$

$$\int_0^{+\infty} x^n e^{-x}\, dx = n! \int \cos\theta\, d\theta = \sin\theta \int_{x^2+y^2 \le R^2} f(x,y)\, dx\, dy = \int_{\theta=0}^{2\pi} \int_{r=0}^{R} f(r\cos\theta, r\sin\theta) r\, dr\, d\theta$$

In non-relativistic wave mechanics, the wave function $\psi(\mathbf{r}, t)$ of a particle satisfies the ***Schrodinger Wave Equation***

$$i\hbar \frac{\partial \psi}{\partial t} = \frac{-\hbar^2}{2m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \psi + V\psi.$$

It is customary to normalize the wave equation by demanding that

$$\iiint_{\mathbf{R}^3} |\psi(\mathbf{r}, 0)|^2\, dx\, dy\, dz = 1$$

A simple calculation using the ***Schrodinger Wave Equation*** shows that

$$\frac{d}{dt} \iiint_{\mathbf{R}^3} |\psi(\mathbf{r}, t)|^2\, dx\, dy\, dz = 0$$

and hence

$$\iiint_{\mathbf{R}^3} |\psi(\mathbf{r}, t)|^2\, dx\, dy\, dz = 1$$

for all times t. If we normalize the wave function in this way then, for any (measurable) subset V of $\mathbf{R}^3$
and time t,

$$\iiint_V |\psi(\mathbf{r}, t)|^2 \, dx \, dy \, dz$$

represents the probability that the particle is to be found within the region V at time t.